

周报

本周工作主要如下：

1. 修改之前将多个位置相同的基站合并成一个基站的做法，把每个基站看成独立的基站进行计算。

➤ 之前的做法：

1) 把基站表格中位置相同的基站合并成一个基站，然后根据合并结果重新对基站进行编号（用 site 标识）。基站表格中的基站由 lac 和 cell 唯一标识，合并后的基站就由编号唯一标识。其中，对于基站表格中有些基站只有 lac 和 cell 值，但没有位置信息的基站，将其 site 编码为-1。

2) 将用户轨迹中，每个记录中的 lac 和 cell 也先都转换成 site 编号。对于轨迹中，有些记录的 lac 和 cell 值在基站表格中找不到的情况，将这些记录的 site 也都统一标记为-1。

3) 在进行数据处理时把轨迹数据中 site=-1 的记录看成无效记录处理

➤ 导致的问题：按照何老师在邮件中的希望的那样，5 天中每个小时的总体都不变。就需要抽取出 5 天中不存在无效点的轨迹。但是，据我之前统计，整个湖州市这样的轨迹只有几百个，这显然过少。

➤ 对问题的思考：

1) 何老师在邮件中提到“只要确定了每个人在家时可能联系的基站，那么，只要这个用户离开了那几个基站，那就是出门了；这个跟基站的位置没有关系”。联系我们的做法，对统计在家和不在家的情况，只要 site 的值不等于判定为家的那个 site 值那就说明是外出，否则就在家。但是对于 site 值为-1 的基站我们不知道具体位置是哪，即我们无法判定它的位置是否跟判定为家的 site 对应的那些基站的位置相同，也就无法判定是否在家。

2) 如果我们不合并基站，那么家就会判定为一个有 lac 和 cell 标记的基站。这样一来根据用户轨迹判断用户是否在家的时候，只要轨迹中记录的 lac 和 cell 值不等于“家”的 lac 和 cell 值，那么不管我们是否知道该基站的位置，都可以判定其是外出。

3) 顾虑：之前之所以会选择合并基站，是因为发现用户轨迹中存在很多 ping-ping effect。觉得同一个位置的基站之间有信号干扰的可能性比较

大。这样修改之后 ping-ping effect 的情况可能会增多，对于居住地判定以及判定一段时间内是否在家，都会有很大的干扰。

➤ 具体做法

1. 统计满足要求的用户轨迹数据

- 方法：忽略“3）”中的顾虑，按照“2）”中的想法，统计了一下轨迹时间跨 2013-12-18 00:00:00 到 2013-12-22:00、且在这两个时间点之间不存在 lac 和 cell 为无效数据的记录（有的用户轨迹中的某些记录的 lac 或者 cell 值为'\N'）、并且可以判定居住的用户的数目。
- 结果：num 代表上一步统计的每个城市用户数，total 代表每个城市所有的轨迹数

	湖州	嘉兴	宁波	台州	温州	舟山
num	393334	476816	1294594	665826	1093595	127929
total	3702388	7246613	8509509	5800685	8738083	783953

2. 进一步延伸时间跨度，因为原始数据时间范围是 2013-12-17 0:00 点到 2013-12-23 0:00。统计时间跨度满足各个时间点（如 2013-12-17 22:00 到 2013-12-22 18:00）的轨迹数目。

- 结果：下面列出了湖州和嘉兴两个城市的统计结果。图中行代表其实时间点（2013-12-17 1:00 到 2013-12-18 0:00），列代表结束时间点（2013-12-22 0:00 到 2013-12-22 23:00）

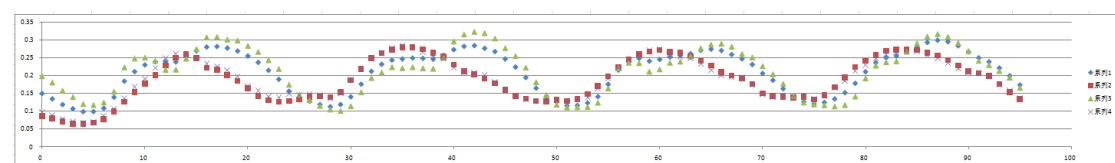
A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y
湖州	22.0	22.1	22.2	22.3	22.4	22.5	22.6	22.7	22.8	22.9	22.10	22.11	22.12	22.13	22.14	22.15	22.16	22.17	22.18	22.19	22.20	22.21	22.22	22.23
17.1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17.2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17.3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17.4	26568	25484	24034	21446	15171	4907	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17.5	78747	75313	70212	62487	40939	13887	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17.6	149138	142396	131008	117098	73935	23870	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17.7	182087	173706	159686	141032	88041	28272	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17.8	193385	184227	169170	149422	93295	29965	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17.9	197251	187920	172690	152683	95402	30568	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17.10	204209	194310	178309	157097	97555	31251	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17.11	218140	207396	190095	167210	103457	32961	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17.12	221174	210201	192604	169477	104925	33462	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17.13	224540	213411	195613	172150	106521	33902	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17.14	228614	217300	199273	175558	108820	34586	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17.15	233108	221637	203400	179382	111216	35504	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17.16	235543	223981	205618	181424	112418	35896	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17.17	237267	225594	207054	182628	113125	36124	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17.18	240224	228285	209463	184683	114278	36479	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17.19	242989	230889	211531	186310	115133	36731	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17.20	264082	251781	232132	206487	134492	55498	18421	18335	18228	18111	17844	17374	17050	16819	16698	16378	15964	15231	14330	12842	8993	2940	0	0
17.21	304837	292198	271697	245173	171613	90884	53235	53008	52690	52376	51526	50392	49466	48879	48561	47554	46477	44180	41155	36698	23976	8213	0	0
17.22	358937	345913	324573	296916	221466	139356	101153	100752	100144	99596	98105	96040	94517	93410	92843	90781	88552	84147	77263	69219	43629	14147	0	0
17.23	383187	369985	348430	320506	244525	161932	123488	123003	122244	121579	119820	117431	115656	114392	113719	111245	108166	102742	94221	83417	52305	16894	0	0
18.0	393334	380010	358280	330141	253732	170777	132162	131638	130817	130103	128163	125594	123688	122343	121637	118889	115471	109496	100282	88785	55740	17990	0	0

嘉兴	22.0	22.1	22.2	22.3	22.4	22.5	22.6	22.7	22.8	22.9	22.10	22.11	22.12	22.13	22.14	22.15	22.16	22.17	22.18	22.19	22.20	22.21	22.22	22.23
17.1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17.2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17.3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17.4	27561	27284	25799	23950	17101	6741	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17.5	103632	102659	97502	89574	68585	22311	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17.6	145186	143780	135090	124058	92755	30889	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17.7	175936	174199	163234	149280	112117	36421	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17.8	222820	220650	206246	189606	141386	47378	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17.9	254856	252372	236054	216774	161644	52502	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17.10	284679	281931	264173	242566	178398	58431	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17.11	307249	304311	284939	261558	192413	62000	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17.12	320728	317650	297606	273184	200255	65264	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17.13	331313	328130	307619	282278	206958	67176	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17.14	344210	340903	319228	293044	214909	70293	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17.15	354338	350829	328571	301414	220819	71905	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17.16	358874	355259	332658	305112	223351	72760	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17.17	361285	357586	334822	307072	224780	73228	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17.18	365112	361340	337826	309800	226761	73875	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17.19	366993	363171	339590	311433	227975	74203	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17.20	381161	377209	353456	325147	241129	86551	11926	11880	11813	11727	11618	11523	11360	11214	11033	10787	10567	10434	9856	9183	6212	2384	0	0
17.21	417046	412872	388874	360291	275607	119867	44729	44539	44334	44018	43679	43305	42821	42228	41602	40486	39705	39257	37224	34183	25679	7991	0	0
17.22	439527	435208	410836	381956	296462	139515	63581	63316	63019	62601	62141	61642	60992	60213	59185	57494	56179	55572	52047	47843	35033	11377	0	0
17.23	454862	450422	425903	396907	311061	153504	77294	76991	76640	76184	75636	75074	74318	73287	72052	69848	68188	67424	62908	57589	42577	13441	0	0
18.0	476816	472286	447596	418469	332330	174378	97922	97552	97123	96558	95908	95246	94530	93166	91742	89069	87035	86105	80078	73673	54014	17814	0	0

3. 抽取了时间跨度满足 2013-12-17 22:00 到 2013-12-22 18:00 重新按之前时间段计算外出用户比例，数据还没给陈吟吟同学，进行回归。下面是湖州的计算结果，其他城市的暂不列出。

城市	时间	在家人数	在家人数比例	外出人数比例
湖州	2013-12-17 22:00--2013-12-18 06:00	70783	0.916142476	0.083857524
湖州	2013-12-18 06:00--2013-12-18 10:00	68453	0.885985349	0.114014651
湖州	2013-12-18 10:00--2013-12-18 14:00	60753	0.786324454	0.213675546
湖州	2013-12-18 14:00--2013-12-18 18:00	58997	0.763596593	0.236403407
湖州	2013-12-18 18:00--2013-12-18 22:00	63901	0.827068934	0.172931066
湖州	2013-12-18 22:00--2013-12-19 06:00	66715	0.863490461	0.136509539
湖州	2013-12-19 06:00--2013-12-19 10:00	59592	0.771297662	0.228702338
湖州	2013-12-19 10:00--2013-12-19 14:00	55966	0.724366441	0.275633559
湖州	2013-12-19 14:00--2013-12-19 18:00	58780	0.760787968	0.239212032
湖州	2013-12-19 18:00--2013-12-19 22:00	63127	0.817051073	0.182948927
湖州	2013-12-19 22:00--2013-12-20 06:00	66884	0.865677824	0.134322176
湖州	2013-12-20 06:00--2013-12-20 10:00	61140	0.791333385	0.208666615
湖州	2013-12-20 10:00--2013-12-20 14:00	56727	0.734216044	0.265783956
湖州	2013-12-20 14:00--2013-12-20 18:00	58241	0.753811706	0.246188294
湖州	2013-12-20 18:00--2013-12-20 22:00	62250	0.805700085	0.194299915
湖州	2013-12-20 22:00--2013-12-21 06:00	66099	0.85551759	0.14448241
湖州	2013-12-21 06:00--2013-12-21 10:00	59593	0.771310605	0.228689395
湖州	2013-12-21 10:00--2013-12-21 14:00	56335	0.729142399	0.270857601
湖州	2013-12-21 14:00--2013-12-21 18:00	58128	0.75234915	0.24765085
湖州	2013-12-21 18:00--2013-12-21 22:00	61955	0.801881908	0.198118092
湖州	2013-12-21 22:00--2013-12-22 06:00	65972	0.853873832	0.146126168
湖州	2013-12-22 06:00--2013-12-22 10:00	57064	0.738577826	0.261422174
湖州	2013-12-22 10:00--2013-12-22 14:00	56067	0.725673682	0.274326318
湖州	2013-12-22 14:00--2013-12-22 18:00	60475	0.782726308	0.217273692

4. 分别抽取湖州市满足时间跨度 2013-12-18 0:00 到 2013-12-22 0:00、2013-12-17 22:00 到 2013-12-22 18:00、2013-12-17 04:00 到 2013-12-22 4:00、2013-12-17 20:00 到 2013-12-22 20:00 的用户分别对应下图中的系列 1 到系列 4，进行每小时外出人数比例的计算，发现有很大的差距。图中横轴分别对应 2013-12-18 0:00 到 2013-12-21 23:00 的每个小时。



➤ 存在的问题

- 1) pm2.5 和天气数据的小时数据还没有查到
- 2) 对于“4”中存在的问题还没能找出问题，代码是有单步调试过的。